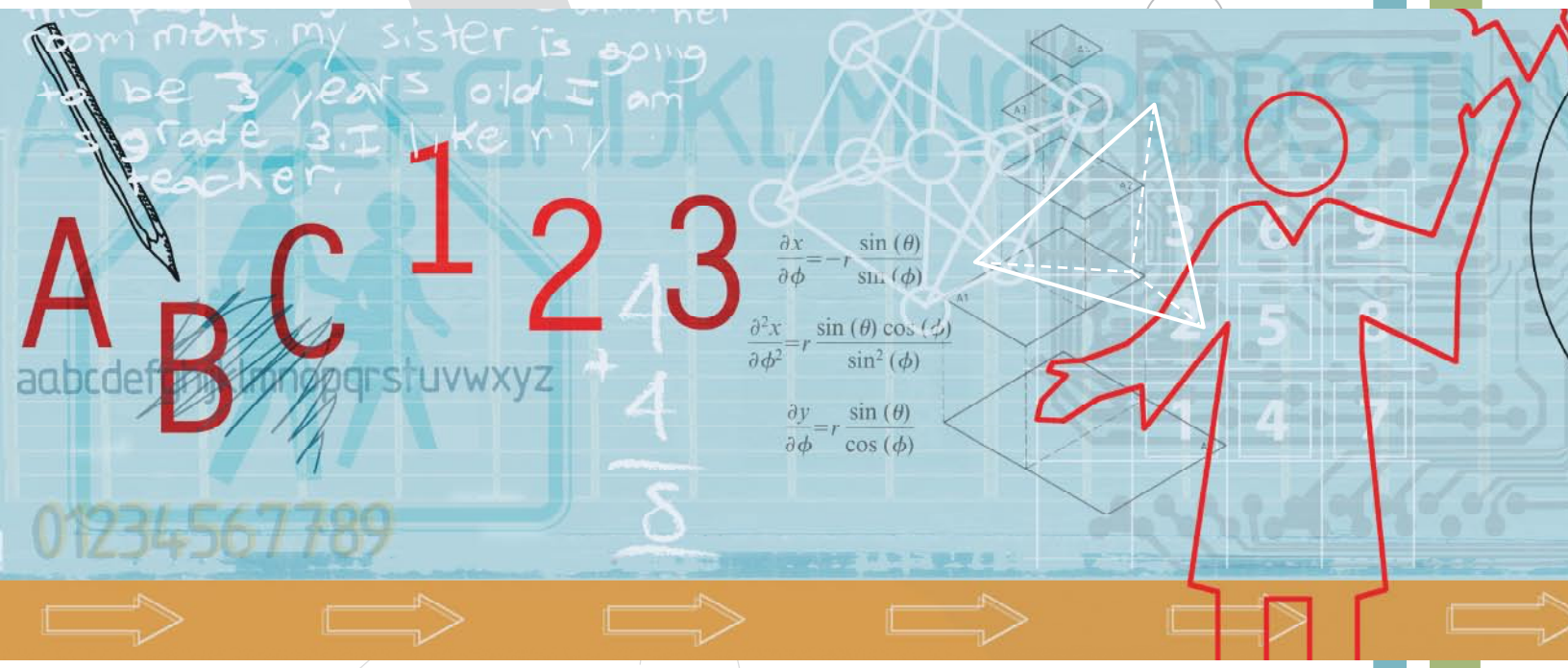
ON BEHALF OF THE PITON FOUNDATION
and the DONNELL-KAY FOUNDATION

TESTING IN COLORADO:

Time, Cost and Purpose



PREPARED BY:

Amy Berk Anderson
Amanda Brown
Bob Palaich

Augenblick, Palaich and Associates

August 2007

THE PITON FOUNDATION

THE COMMUNITY INVESTMENT DIVISION OF GARY-WILLIAMS ENERGY CORPORATION

DONNELL-KAY FOUNDATION

August 21, 2007

Dear Reader:

The report that accompanies this letter, ***Testing in Colorado: Time, Cost, and Purpose***, was commissioned by the Piton Foundation and the Donnell-Kay Foundation to document the types of assessments that Colorado students take, and the time and costs associated with this testing. Given the strong sentiments that surround testing, it is important to analyze the testing landscape in Colorado. Because this issue stirs so much debate, this report seeks to provide research-based data and information toward a more productive and informed policy discussion around testing an assessment.

The report presents a series of facts about testing in the state, followed by some questions that surfaced during Augenblick, Palaich and Associates' (APA's) research, including:

- How well do our existing assessments measure what kids need to know in order to be successful in college and the workforce?
- Is the current amount of time and associated expense for testing adequate? Or, conversely, are we spending too much time and money testing kids? What is the appropriate balance (in time and money) between assessing what kids need to know through summative and formative assessments?
- How should assessment results be shared and how can we ensure that results are used in meaningful ways?

This is one of several reports that Donnell-Kay and Piton will release this year that have been researched and written specifically for policy-makers and members of the Governor's P-20 council.

We welcome your feedback and hope you find this report to be helpful in your work.

Sincerely,

Tony Lewis
Executive Director, Donnell-Kay Foundation

Van Schoales
Program Officer, Education, Piton Foundation

Table of Contents

Executive Summary

Objective

Colorado students are engaged in various types of testing each year. The purpose and quality of such testing, the time spent taking and administering tests, and the usefulness of results is an ongoing topic of discussion among educators and policy-makers. In order to get a better sense about the testing occurring in Colorado, the Piton Foundation and Donnell-Kay Foundation commissioned Augenblick, Palaich and Associates (APA) to document the types of assessments that students in our state are taking annually, as well as the associated time and expense of such tests.

Methodology

In order to gather this information, APA interviewed officials and collected data from the Colorado Department of Education, three school districts of varying sizes and student characteristics (Denver, Littleton, Ouray), and others in the field who are knowledgeable about assessment at the state and national levels. Information presented in this section summarizes what we learned from these interviews and the other data collected. Section II provides detailed information about the various types of assessments that students in Colorado are engaged in on an annual basis.

Findings

Colorado schools engage primarily in six types of tests:

- Summative assessments like CSAP, required by state and used to assess progress in meeting state standards;
- Interim/benchmarking assessments, used by districts to measure student growth over the course of the year;
- Formative assessments, used by schools and teachers as part of instruction;
- Diagnostic testing, such as testing for English-language learners, special education and gifted students;
- College entrance exams, such as the ACT, used to predict success in college; and
- National tests, like the National Assessment of Educational Progress, used to measure students across country.

Most Colorado students spend at least 12 hours each year on academic assessments. On average, 5th and 8th graders spend just under 2% of total school hours on testing. The CSAP is the most time-consuming. These figures do not represent preparation time. At the state and local levels, Colorado spends an estimated $54.59 per year per pupil across grades 2-11 on selected summative, formative and interim tests.

An Overview of Testing: Types, Time, and Costs

Types of Tests

Many types of assessments are used by Colorado schools. Some are required by the state while others are chosen or developed at the district or school levels. An in-depth discussion of each of the various types of tests is presented in the pages that follow. A summary is presented below.

Summative Assessments (required by the state)— Colorado Student Assessment Program (CSAP) and Colorado Student Assessment Program Alternate (CSAP A) are both summative assessments required by the state. The CSAP is designed to assess students' progress in meeting state standards in a given year. Tests are conducted in the spring as a means of "summing up" how well students did that year in school which is why these tests are referred to as summative assessments. CSAP results are tied to the state's accountability program.

Interim/Benchmarking Assessments (administered by school districts)—Most districts use some type of interim or benchmarking assessments to measure student growth towards meeting standards over the course of a school year. Unlike the CSAP, results from these assessments are available during the school year and can be used to guide instructional change based on how kids perform.

Formative Assessments—Formative assessment is "defined as assessment carried out during the instructional process for the purpose of improving teaching or learning….frequent, interactive assessments of student progress and understanding to identify learning needs and adjust teaching appropriately."

Other Academic Assessments (administered by school districts)—Colorado school districts are required to administer assessments to students in the primary grades (K-3) on basic literacy. Additionally, districts may choose, at their discretion, to administer other assessments to all students or a group of students by grade. College-bound students are often interested in taking Advanced Placement or International Baccalaureate exams, but usually bear costs.

Diagnostic Testing for Students with Special Needs— Districts use various assessments to identify special education students and English language learners and to figure out how best to serve their needs.

College Entrance Exams—All 11th grade students in the state are required to take the Colorado ACT, an exam that is accepted at most universities for admission.

National Testing—A sampling of students in grades 4,8,12 across the state participate annually in the National Assessment of Educational Progress (NAEP). The NAEP is regarded as the Nation's Report Card and measures how students are doing across the nation (relative to each other and students in other countries).

An Overview of Testing: Types, Time, and Costs

Time Commitment to Testing

Most Colorado students are spending at least a dozen hours per year on CSAP, NAEP, ACT, and selected other district academic assessments. 8th grade students spend the most time on tests (17.5 hours per pupil per year on average) and K, 1, and 11th grade students the least (2 hours per pupil per year for K-1 and 3 hours per pupil per year for 11th grade students). On average, 5th and 8th graders are spending just under 2% of the total hours they are required to be in school per year on testing. The CSAP tests are the most time consuming of all of the tests examined. Table 1 provides more detailed information about the hours spent on selected tests per year by grade.

Not included in these estimates are the hours spent testing students with special needs or the time spent by adults developing, administering, and analyzing tests. The amount of time spent in test preparation often is greatest for students in poor performing schools—time that would otherwise be spent on instruction.

Table 1: Hours spent on selected tests per year

	K	1	2	3	4	5	6	7	8	9	10	11	12	Total Hours
CSAP				6	9	12	9	9	12	9	12			78
NAEP					1.5				1.5				1.5	4.5
MAP			4	4	4	4	4	4	4	4				32
Basic Literacy	2	2	2	2										8
CO ACT												3		3
Total per pupil	**2**	**2**	**6**	**12**	**14.5**	**16**	**13**	**13**	**17.5**	**13**	**12**	**3**	**1.5**	**125.5**

Cost of Tests

At the state and local levels, Colorado spends an estimated $54.59 per year per pupil (across grades 2-11) on selected summative, formative, and interim tests. The grades with the least amount spent on testing were 2nd and 11th, with the amount spent in the other grades about the same (approximately $64 per pupil). Tables 2 and 3 provide aggregated and estimated per pupil costs for selected tests.

Table 2: Cost (amount spent on selected tests annually)

	Total amount spent annually	Approximate cost per pupil	Required by
CSAP	In 2005, nearly $25 million was spent on CSAP in CO ($9 million came from the US DOE).	$52.43 (grades 3-10)	State, Federal
NAEP	Approximately $93 million was budgeted for the NAEP (nationally) in 2006 and over 930,000 students took the test (total from grades 4,8,12)	$100 (no cost to state)	Federal Government
MAP	$2300 (Ouray), $95,000 (Littleton)	$10-14	District/school
Successmaker	$1100 per license (one license serves approx. 10 kids per week). An average school has 300 kids per week using 30 licenses at a cost of $33,000 per year	$110	District/school
DPS Benchmark	$900,000	$12	District
Basic Literacy	Set-up licensing fee (DPS spent $350/ K-5 student). After this licensing fee, subsequent costs are for personnel time spent administering and scoring the test.	Personnel-related	District
CO Act	$1.6 million (approximately)	$29	State

This chart does not include all of the tests that students may need to take but only those that all or most must take (not included are those unique to a given district or those to assess ELL or Special Education needs).

Table 3: Statewide cost estimates of selected tests per grade

	2	3	4	5	6	7	8	9	10	11	Total Estimated Cost
CSAP		$2,998,944	$2,996,427	$2,994,277	$3,023,848	$3,094,523	$3,143,074	$3,347,184	$3,145,485		$24,743,762
CO ACT										$1,600,000	$1,600,000
MAP	$723,696	$709,512	$694,512	$694,869	$694,116	$700,440	$713,316	$775,836			$5,706,288
Total	$723,696	$3,708,456	$3,690,939	$3,689,137	$3,717,964	$3,794,963	$3,856,390	$4,123,020	$3,145,485	$1,600,000	$32,050,050
Total per pupil	$12.33	$64.83	$64.58	$64.60	$64.47	$64.30	$64.33	$64.58	$52.43	$29.43	$54.59

Conclusion

The overall purpose of this document is to examine testing practices as they exist today in Colorado. However, various questions and recommendations about testing surfaced as APA conducted this research. Below, we share some of the questions that have been raised about assessment that warrant further attention as Colorado considers how to improve its assessment system.

How well do our existing assessments measure what kids need to know in order to be successful in college and the workforce?

Our nation's system of testing has been criticized as inadequate in preparing kids for college and the workforce. Such criticism has been brought by organizations such as the National Center on Education and the Economy (NCEE), Achieve, and the Partnership for 21st Century Skills.

These groups question whether current assessments measure the kinds of skills that employers seek or the advanced concepts and skills that college professors indicate are important. Because of these deficiencies, states like Maine, Michigan and Kentucky are augmenting their college entrance exams and state assessments in order to better capture students' understanding of necessary skills.

A recent article from *Education Week*, authored by leaders from the Partnership for 21st Century Skills, argues for a system of assessment that measures problem-solving and reasoning strategies that are difficult to capture in the type of multiple choice tests currently given in our country. In its recently released book, *Tough Choices or Tough Times*, NCEE also stresses the importance of an education system that nurtures and assesses the creative side of learning. Other countries, like Singapore, are shifting their educational systems and assessments in order to better prepare students for the future—one where strong analytical, entrepreneurial, and dynamic skills are critical. Tests scores from the Program for International Student Assessment (PISA) show 15 yr-old students in the U.S. lagging behind those in Europe and Asia in problem-solving skills in mathematics and science.

Colorado leaders have partnered with Achieve, Inc. to examine our assessment system's alignment with critical skills needed for post-secondary education and beyond. This type of analysis, as well as a further examination of the recommendations from places like NCEE and the Partnership for 21st Century Skills and their applicability to Colorado students, is an essential next step to ensure that Colorado students leave high school prepared to succeed in college and life.

Is the current amount of time and associated expense for testing adequate? Or, conversely, are we spending too much time and money testing kids? What is the appropriate balance (in time and money) between assessing what kids need to know through summative and formative assessments? Most students in our state are spending a dozen or more hours per year on various assessments. Academics argue that this time would be best spent on high quality formative assessments because research indicates that the use of such assessments can have a more positive impact on achievement than many instructional interventions.

While the CSAP has been touted by the U.S. Department of Education as one of the best state assessments in the nation, do we know if it adequately prepares kids for skills needed in college? During APA's interviews, questions arose as to whether it was necessary to test kids every year with the CSAP or spread it out to every 2-3 years and use targeted formative or interim assessments in off years (or essay-type international assessments) to assess students' mastery of skills beyond what can be measured in a standard multiple choice exam. APA encourages Colorado's policy-makers and educators to closely evaluate the effectiveness and use of current assessments. Do they test what kids really need to know and, if not, what is needed to augment our current system or to make it more efficient?

How should assessment results be shared and how can we ensure that results are used in meaningful ways?

A significant issue that surfaced in our interviews was how to use assessment data and with whom to share the data. In many cases, parents do not see their children's formative assessment data. And, when data are shared with parents, (including CSAP results) little to no explanation accompanies the results. Therefore, parents either do not understand the information or need to do their own research in order to assess its meaning. Furthermore, students take these tests yet rarely receive any feedback on how they did (e.g., what they knew, where they struggled, how they compared with their peers, etc.). How can schools, districts, and the state

better inform students and their families about assessments—not just the actual scores kids receive but the purpose, importance, and results of the assessments being used?

A second issue raised concerned a lack of capacity among teachers, principals, and district administrators about what to do with assessment results. And, in many cases, the information provided by testing companies is not detailed enough to actually help a teacher understand the specific skills students know and don't know within a given area.

A third issue which surfaced in APA's interviews is whether school systems are appropriately collecting and analyzing data across schools. In fact, concerns were voiced as to whether school districts have the capacity to warehouse and analyze their own assessment data, and whether there is a role for the Colorado Department of Education to help in this regard. Questions for Colorado policy-makers to consider include:

- Would a state-held data warehouse allow for better cross-district/statewide comparisons of student learning?
- Would it make sense to have a set of "state-approved" assessments (formative and summative) that districts could opt-into as a means of developing a more uniform set of practices and comparable data across the state?

Each of these questions are important and worthy of further attention.

Supplemental Information about the Types of Tests Colorado Students Take

The following pages provide detailed information about the various types of tests that Colorado students take each year:

- Summative Assessments
- Interim/Benchmarking Assessments
- Formative Assessments
- Other Academic Assessments
- Diagnostic Assessments
- College Entrance Exams
- National Exams

Included in each description is a description of the test(s), the approximate time commitment annually for students (by grade), the cost of the test, and strengths and challenges of the tests.

Summative Assessments

Description:

Colorado Student Assessment Program (CSAP) and Colorado Student Assessment Program Alternate (CSAP A) are both summative assessments required by the state. All students in grades 3-10, with the exception of those who have significant cognitive disabilities, are required to take the CSAP annually. Those who have these significant disabilities are required to take the CSAP A (about .5% of the student population in CO takes the CSAP A). The CSAP is designed to assess students' progress in meeting state standards in a given year. Tests are conducted in the spring as a means of "summing up" how well students did that year in school which is why these tests are referred to as summative assessments. CSAP results are tied to the state's accountability program.

Time Commitment:

- 3rd grade: 6 hours
- 4th, 6th, 7th, & 9th grades: 9 hours
- 5th, 8th, & 10th grades: 12 hours
- These time estimates do not include time spent on test preparation or time spent by adults related to the CSAP.

Supplemental Information

Cost:
According to the 2007-08 joint budget committee's staff budget briefing on the CO Department of Education's budget, $24,765,344 was spent in fiscal year 2005-06 on the CSAP. The legislative appropriation was lower in 2006 ($21,771,340).

Strengths and Limitations:

- Thorough measure of students' progress in meeting state standards, as recognized by the U.S. Department of Education.
- Allows for comparisons across schools and districts.
- Includes constructed and multiple choice response items, thereby allowing students to express their knowledge in diverse ways.
- Time consuming—students spend more than a dozen hours on CSAP tests each year.
- Results are not immediate (released in August after the school year ends).
- Test is not designed to provide diagnostic information so it is difficult to use results to guide instructional change.
- The Colorado Education Alignment Council formed in 2005 by then Governor Owens is charged with ensuring that Colorado's model content standards are aligned with the expectations of the postsecondary education system and the workforce (and the CSAP exam is designed to measure progress in meeting such alignment). This Council is supposed to share the results of its examination of the alignment starting this Spring (2007) and recommend changes to the standards and CSAP exams by next Spring (2008).

Interim/Benchmarking Assessments

Description:
Most districts use some type of interim or benchmarking assessments to measure student growth towards meeting standards over the course of a school year. Unlike the CSAP, results from these assessments are available during the school year and can be used to guide instructional change based on how kids perform. Most districts use interim/benchmarking assessments a couple of times a year to measure where kids are at the beginning of the year and what they have learned by the end of the year (e.g., MAP, SAT 10). In some cases, districts choose these types of assessments up to four times per year in order to assess growth and to determine if an instructional course of action is needed for the remainder of the year. Results are available relatively quickly (immediately in the case of the computer-based MAP tests, within a few weeks for the SAT 10 exams, and within days for the DPS benchmarking exams).

Time Commitment:
DPS developed its own benchmarking exams which are given three times a year to students in grades 3-10, and once at the end of the 2nd grade. Each test is one hour long.

MAP tests are administered twice per year in the Fall and Spring (in some cases, a district or school may choose to do a third or fourth test during the year). In general, the Fall tests provide a benchmark of student knowledge at the beginning of the school year. The Spring tests measure how much students have grown (learned) over the course of the school year. Students generally take MAP tests in math and reading. Each test is one-hour long so students spend two hours in the Fall and two hours in the Spring taking these tests. In the two districts we spoke with, all students in grades 2-9 take the MAP tests each year.

Cost:
DPS spends approximately $900,000 per year on this test. Costs vary for the MAP test based on the number of students taking the exam. The smallest district APA surveyed spends $2300 per year on MAP testing (or about $14/student, grades 2-9). The moderate size district spends $95,000 annually (or about $10/student, grades 2-9).

Strengths & Limitations of these Tests:

- In the case of the DPS benchmarks, the district has control over the content of the tests, and teachers can have feedback within days that can be used to improve classroom instruction.
- Computer-based tests like MAP are adaptive, meaning that they dynamically adjust to each student's performance level. As a student answers a test question on a computer, the program immediately analyzes the student's response, and based on how well the student has answered previous questions, selects a question of appropriate difficulty to display next.

- DPS has spent a significant amount of staff time to develop their tests, fine-tuning as they go. Because DPS is the only district using this particular test, students in DPS can't be compared with peers nationally (or in other CO districts) in the same way that students who take other types of formative, norm-referenced exams can (e.g., SAT 10).
- In the case of MAP tests, some districts indicated that the analysis is not as detailed as they were led to believe it would be when they started out using the program. MAP tests do a good job telling teachers how well students are likely to do on the CSAP but they don't tell them exactly which concepts students are understanding (or not understanding).
- Interim assessments have been criticized for measuring low-level skills via multiple choice responses. For some areas this is fine, but for higher level conceptual skills and writing, it is inadequate.

Formative Assessments

Description:

While some testing companies package interim/benchmarking assessments as formative assessments, testing experts argue that a true formative assessment is quite different than an interim assessment like the MAP tests. Formative assessment is "defined as assessment carried out during the instructional process for the purpose of improving teaching or learning….frequent, interactive assessments of student progress and understanding to identify learning needs and adjust teaching appropriately."

The primary difference between formative and interim assessments is that formative assessments are given frequently to measure students' mastery of specific skills that a teacher is teaching in a classroom. Results are used immediately to alter instruction (as opposed to interim assessments which are used periodically during the year to evaluate students' mastery of key outcomes tied to standards).

Some test developers are moving towards electronically administered formative assessments that provide a level of detail about specific skills being taught in the classroom that extend beyond the type of information provided in interim assessments (e.g., Successmaker, Dynamic Indicators of Basic Early Literacy Skills-DIBELS). Like the MAP or SAT 10 tests, these assessments can analyze student progress over the course of the year via benchmarking exams. However, in addition to the benchmarking, frequent check-ins are conducted with students throughout the year (45-60 min. per week on average). Results are available immediately and used by teachers to differentiate and guide instruction in various content and skill areas based on the class-wide and individual student-level results. These "just in time" formative assessment are adaptive and computer-based which means that they can adjust to each student's performance level while the student is taking the exam.

Time Commitment:

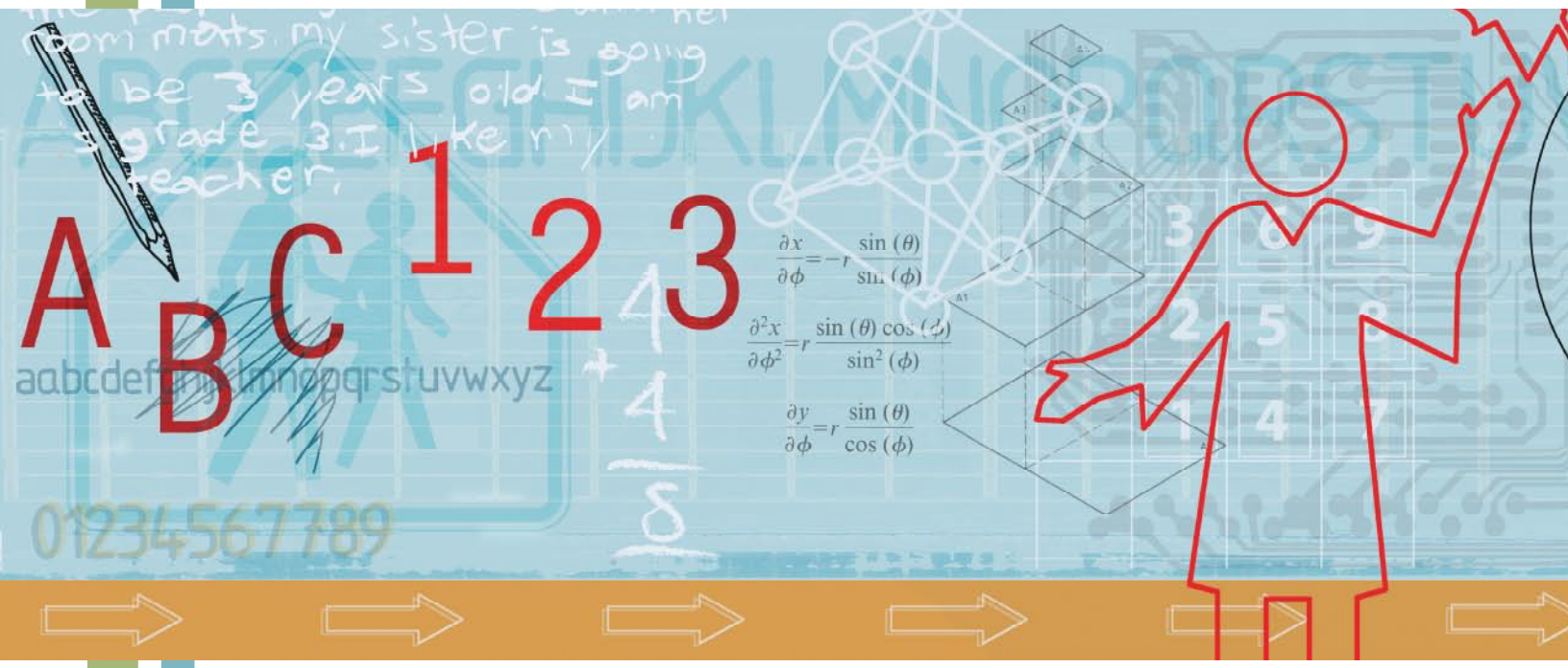
Successmaker is being used in approximately 35-40 school districts in Colorado. In some cases only a school or two within the district is using the program while in others, like Pueblo 60, most elementary schools are using the program. On average, students spend 45-60 minutes per week (15-20 min. per day, 3 times per week) on the computer using the Successmaker program.

Cost:

Schools license the Successmaker program at $1100 per license and approximately 10-15 students use one license per day (one license covers various content areas including reading, writing, spelling, math, ESL, science, and Algebra). The per pupil costs is approximately $110 for the Successmaker program.

Strengths & Limitations of these Tests:

- The data alone is not enough. School leaders and teachers need to be trained on how to use the information provided to them (e.g., how to change how and what they teach based on what they see in the formative data results, how to read the data, etc.
- In some instances, students using Successmaker have moved ahead of what the teacher has covered in classroom instruction which can cause some frustration among teachers. However, it also allows students to progress at their own pace and master key concepts more quickly than they might otherwise. The challenge for teachers is to figure out how to differentiate instruction based on the varying needs and skill levels of their students.
- The use of the term "formative assessments" is a subject of considerable debate in the assessment community. While several companies market their assessments as formative assessments (e.g., MAP), testing experts argue that such tests do not fit the true definition of the term because they fail to provide appropriate information to help teachers understand students understanding of key concepts.



THE PITON FOUNDATION
THE COMMUNITY INVESTMENT DIVISION OF GARY-WILLIAMS ENERGY CORPORATION

DONNELL-KAY FOUNDATION

Acknowledgements

APA would like to thank Dr. Lorrie Shepard of the University of Colorado for her review of earlier drafts of this report and for her valuable input on testing that informed important changes to the report resulting in this final version.